



Машинное обучение для решения исследовательских и инженерных задач в науках о Земле

Михаил Криницкий

krinitsky.ma@phystech.edu

К.Т.Н., Н.С.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга
климатических изменений (ЛВОАМКИ)



[Resources policy](#)

[Homeworks policy](#)

Михаил Криницкий

krinitsky.ma@phystech.edu

К.Т.Н., Н.С.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга
климатических изменений (ЛВОАМКИ)

Общий принцип обучения по прецедентам (оптимизация функции ошибки)

$x \in \mathbb{X}$ — объекты, objects

$y \in \mathbb{Y}$ — ответы, labels

$\mathcal{F}: \mathbb{X} \rightarrow \mathbb{Y}$ — искомая закономерность

$\mathcal{T}: \{x_i; y_i\}$ — «обучающая выборка»
(прецеденты), train dataset

Найти: $\hat{\mathcal{F}}: \{x_i\} \rightarrow \{y_i\}$

один из способов решения:

$\mathcal{L}(\hat{\mathcal{F}}(x))$ — функционал ошибки
(эмпирического риска, потерь), Loss function

$\hat{y}_i = \hat{\mathcal{F}}(x_i) = f(\vec{p}, x_i)$ — функционально задаваемая зависимость. **Предположение исследователя о виде закономерности.** Иногда задается параметрически, \vec{p} — вектор параметров.

$\mathcal{L} = L(\vec{p}, \mathcal{T})$ — функция ошибки

$$\hat{p} = \underset{\mathbb{P}}{\operatorname{argmin}}(L(\vec{p}, \mathcal{T}))$$

$$\hat{\mathcal{F}} = f(\hat{p}, x)$$

Общий принцип обучения по прецедентам (оптимизация функции ошибки)

$x \in \mathbb{X}$ — объекты, objects

$y \in \mathbb{Y}$ — ответы, labels

$\mathcal{F}: \mathbb{X} \rightarrow \mathbb{Y}$ — искомая закономерность

$\mathcal{T}: \{x_i; y_i\}$ — «обучающая выборка»
(прецеденты), train dataset

Найти: $\hat{\mathcal{F}}: \{x_i\} \rightarrow \{y_i\}$

один из способов решения:

$\mathcal{L}(\hat{\mathcal{F}}(x))$ — функционал ошибки

(эмпирического риска, потерь), Loss function

**Чем руководствоваться при выборе
функции ошибки?**

$\hat{y}_i = \hat{\mathcal{F}}(x_i) = f(\vec{p}, x_i)$ — функционально
задаваемая закономерность. **КАКИЕ бывают функции ошибки?!**

исследователя о виде закономерности. Иногда
задается параметрически, \vec{p} — вектор параметров.

$\mathcal{L} = L(\vec{p}, \mathcal{T})$ — функция ошибки

$$\hat{p} = \underset{\mathbb{P}}{\operatorname{argmin}}(L(\vec{p}, \mathcal{T}))$$

$$\hat{\mathcal{F}} = f(\hat{p}, x)$$

Обучение по прецедентам: вероятностная постановка

принцип максимального правдоподобия maximum likelihood estimation

\vec{x}_i - признаковое описание объектов

\vec{y}_i - признаковое описание ответов

$p(\vec{x}, \vec{y})$ – (искомая,
аппроксимируемая) совместная
плотность распределения событий
на множестве $X \times Y$

$\mathcal{T}: \{\vec{x}_i; \vec{y}_i\}$ — «обучающая выборка»
(прецеденты), train dataset

Предположение!

(x_i, y_i) – выбираются из $p(x, y)$
независимо и случайно

Обучение по прецедентам: вероятностная постановка

принцип максимального правдоподобия maximum likelihood estimation

\mathbf{x}_i - признаковое описание объектов

\mathbf{y}_i - признаковое описание ответов

$p(\mathbf{x}, \mathbf{y})$ – (искомая, аппроксимируемая)
совместная плотность распределения
событий на множестве $X \times Y$

$\phi(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta})$ - модель плотности
распределения, предлагаемая
исследователем

$\mathcal{T}: \{\mathbf{x}_i; \mathbf{y}_i\}$ — «обучающая выборка»
(прецеденты), train dataset

Предположение!

$(\mathbf{x}_i, \mathbf{y}_i)$ – выбираются из $p(\mathbf{x}, \mathbf{y})$
независимо и случайно

Обучение по прецедентам: вероятностная постановка

принцип максимального правдоподобия maximum likelihood estimation

x_i - признаковое описание объектов
 y_i - признаковое описание ответов
 $p(x, y)$ – (искомая, аппроксимируемая)
совместная плотность распределения
событий на множестве $X \times Y$
 $\phi(x, y, \theta)$ - модель плотности
распределения, предлагаемая
исследователем

$\mathcal{T}: \{x_i; y_i\}$ — «обучающая выборка»
(прецеденты), train dataset

Предположение!

(x_i, y_i) – выбираются из $p(x, y)$
независимо и случайно

MLE

$\phi(x_i, y_i, \theta)$ - правдоподобие для одного экземпляра выборки

$L(\{x_i\}, \{y_i\}, \theta) = \prod_{i=1}^N \phi(x_i, y_i, \theta)$ - правдоподобие выборки

$$\theta^* = \underset{\Theta}{\operatorname{argmax}} L(\{x_i\}, \{y_i\}, \theta)$$

Функция потерь определяется видом модели плотности
распределения $\phi(x, y, \theta)$, предложенной исследователем!

Правдоподобие выборки $L(\theta, \mathcal{T})$ – **максимизировать** (в
пространстве параметров Θ)

Функцию потерь $\mathcal{L}(\theta, \mathcal{T})$ – **минимизировать** (в пространстве
параметров Θ)

Обучение по прецедентам. Вероятностная постановка, MLE

Примеры

Линейная регрессия

MSE

$$\phi(x, y, \theta) = \theta x + \epsilon,$$

$$p(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\epsilon^2}{2\sigma^2}}$$

$$\Rightarrow \mathcal{L}(\theta, \mathcal{T}) = \frac{1}{N} \sum_{i=1}^N (y_i - \theta x_i)^2$$

MAE

$$\phi(x, y, \theta) = \theta x + \epsilon,$$

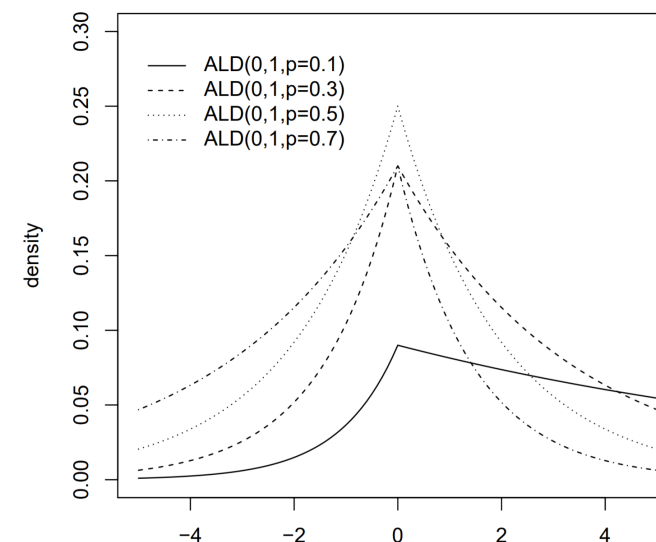
$$p(\epsilon) = \frac{1}{2b} e^{-\frac{|\epsilon|}{b}}$$

$$\Rightarrow \mathcal{L}(\theta, \mathcal{T}) = \frac{1}{N} \sum_{i=1}^N |y_i - \theta x_i|$$

Квантильная регрессия

Bera, Anil & Galvao, Antonio & Montes-Rojas, Gabriel & Park, Sung Y.. (2015). **Asymmetric Laplace Regression: Maximum Likelihood, Maximum Entropy and Quantile Regression**. Journal of Econometric Methods. 10.1515/jem-2014-0018.

Sánchez, B. L., Lachos, H. V., & Labra, V. F. (2013). **Likelihood based inference for quantile regression using the asymmetric Laplace distribution**. Journal of Statistical Computation and Simulation, 81, 1565-1578.



Asymmetric Laplace density. From Sánchez et al.

Обучение по прецедентам. Вероятностная постановка, MLE

Примеры

Логистическая регрессия (logistic regression)

Coming soon!
(its for classification problem)



Решение задач типа ОБУЧЕНИЕ С УЧИТЕЛЕМ

Михаил Криницкий

krinitsky.ma@phystech.edu

К.Т.Н., Н.С.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга
климатических изменений (ЛВОАМКИ)

...

Вечор, ты помнишь, вьюга злилась,
На мутном небе мгла носилась;
Луна, как бледное пятно,
Сквозь тучи мрачные желтела,
И ты печальная сидела —
А нынче... погляди в окно:

Под голубыми небесами
Великолепными коврами,
Блестя на солнце, снег лежит;
Прозрачный лес один чернеет,
И ель сквозь иней зеленеет,
И речка подо льдом блестит.

...

А.С. Пушкин, «Зимнее утро»

...

Буря мглою небо кроет,
Вихри снежные крутя;
То, как зверь, она завоет,
То заплачет, как дитя.
Выпьем, добрая подружка
Бедной юности моей,
Выпьем с горя; где же кружка?
Сердцу будет веселей.

А.С. Пушкин, «Зимний вечер»

...

Было так: Нева, как зверь, стонала,
Серые ломая гребешки,
Колыхались барки у причала,
И царапал стынувшие щеки
Острый дождь, ложась, как плащ широкий,
Над гранитным логовом реки.

...

В. Рождественский, «Октябрьская погода»

О признаковом описании событий (объектов) в геофизике

...

Вечор, ты помнишь, вьюга злилась,
На мутном небе мгла носилась;
Луна, как бледное пятно,
Сквозь тучи мрачные желтела,
И ты печальная сидела —
А нынче... погляди в окно:

Под голубыми небесами
Великолепными коврами,
Блестя на солнце, снег лежит;
Прозрачный лес один чернеет,
И ель сквозь иней зеленеет,
И речка подо льдом блестит.

...

А.С. Пушкин, «Зимнее утро»

...

Буря мглою небо кроет,
Вихри снежные крутя;
То, как зверь, она завоет,
То заплачет, как дитя.
Выпьем, добрая подружка
Бедной юности моей,
Выпьем с горя; где же кружка?
Сердцу будет веселей.

А.С. Пушкин, «Зимний вечер»

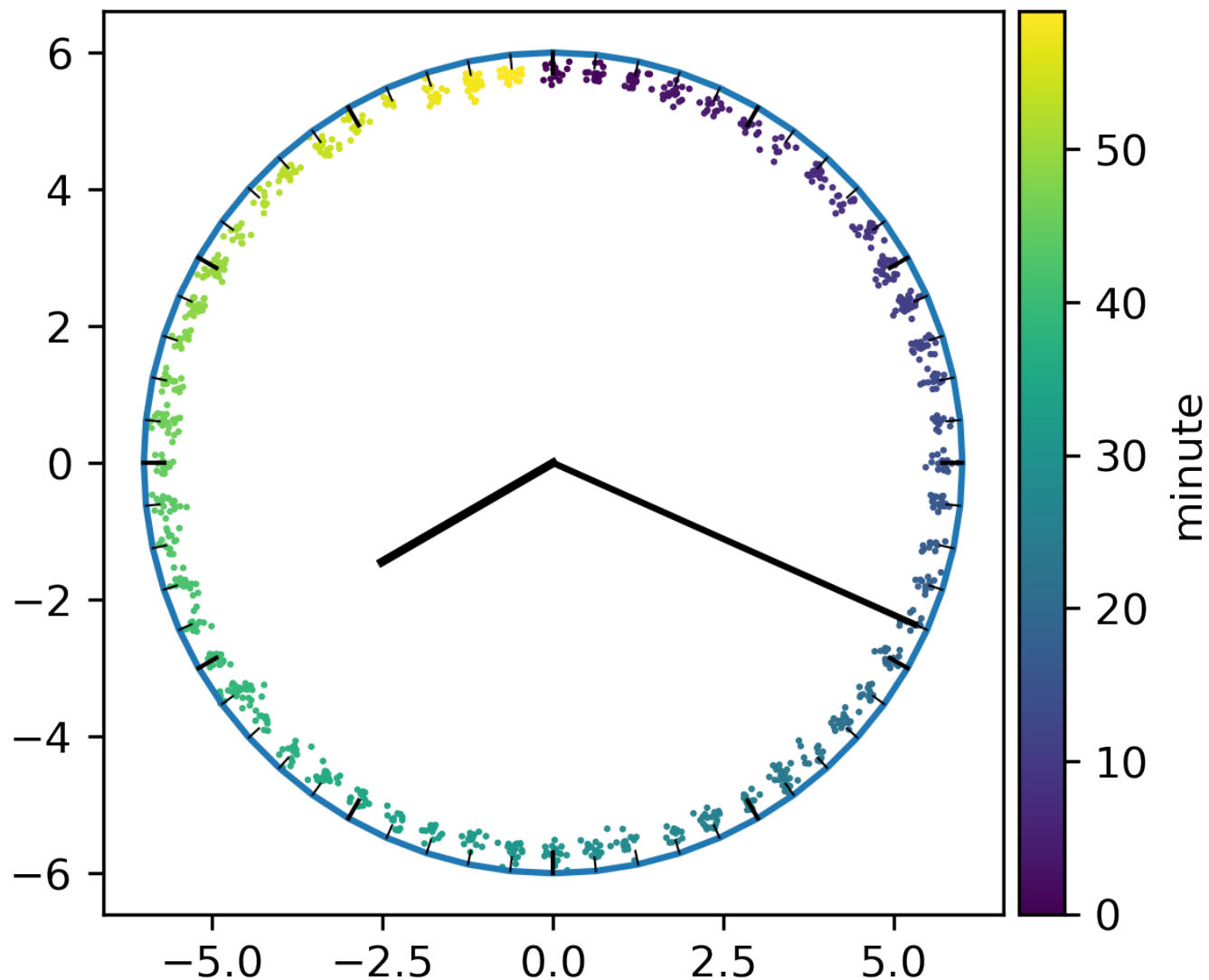
...

Было так: Нева, как зверь, стонала,
Серые ломая гребешки,
Колыхались барки у причала,
И царапал стынувшие щеки
Острый дождь, ложась, как плащ широкий,
Над гранитным логовом реки.

...

В. Рождественский, «Октябрьская погода»

Решение задачи восстановления регрессии: ПРИМЕР



Синтетическая задача, “toy problem”

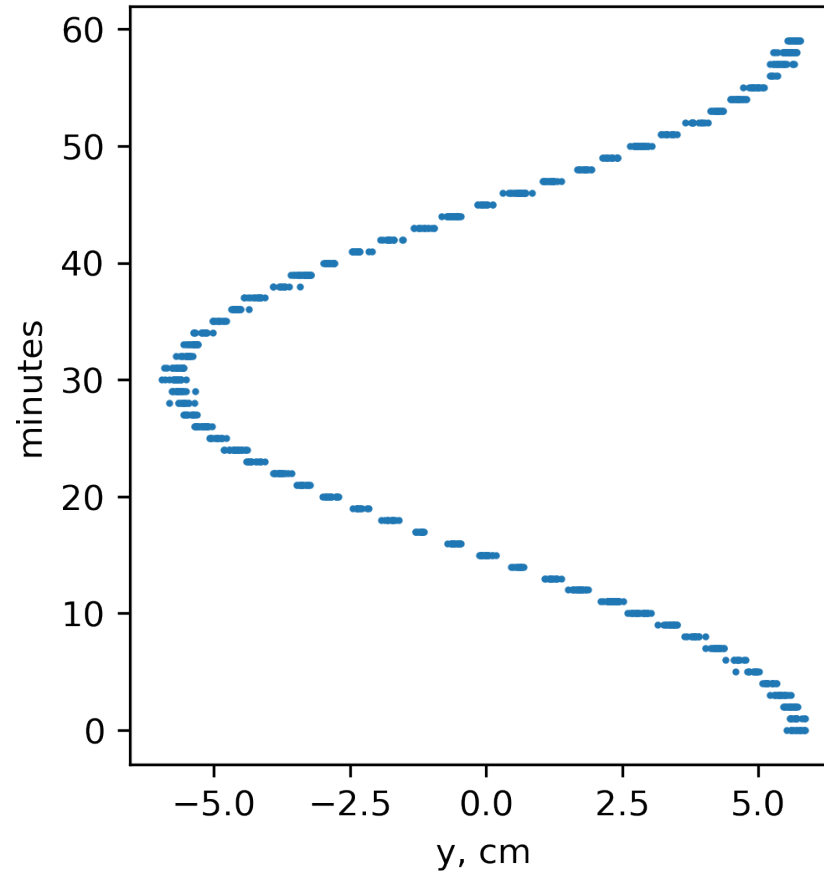
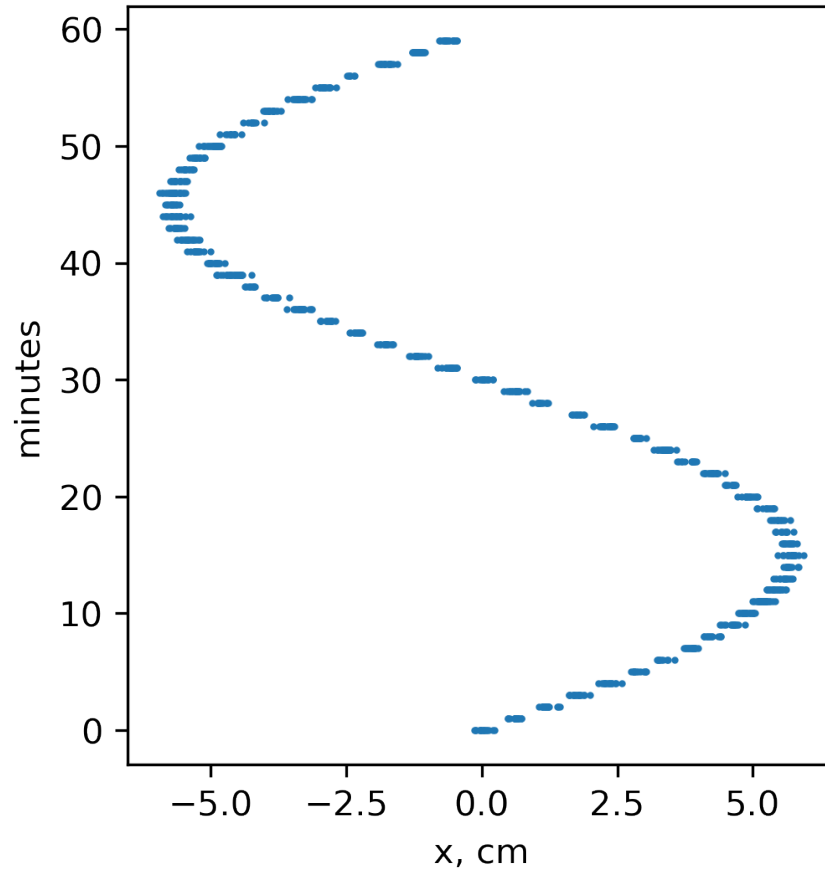
События x_i : наблюдения циферблата часов

Признаковое описание событий \vec{x}_i : координаты
конца минутной стрелки

Целевая переменная m_i : минутная компонента
времени

Решение задачи восстановления регрессии: ПРИМЕР

Исследование данных: визуализация, поиск структуры



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем очень слабую модель

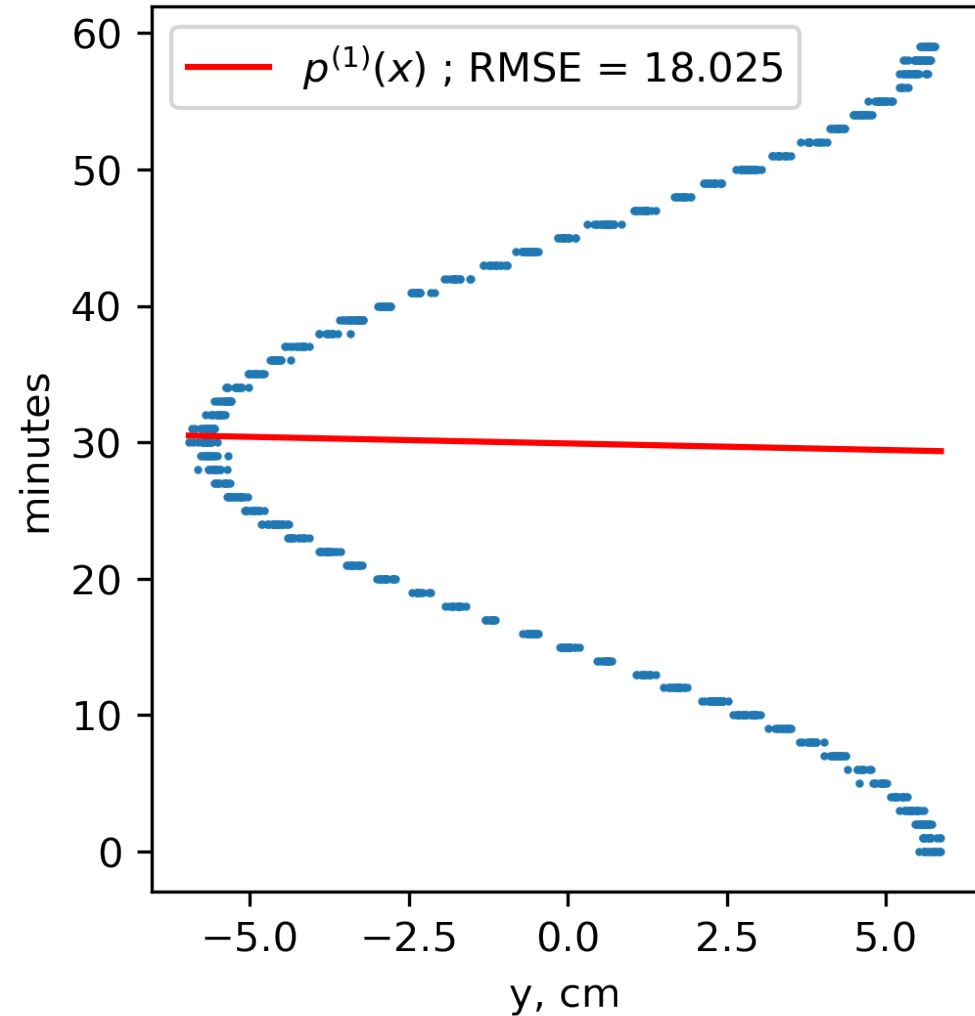
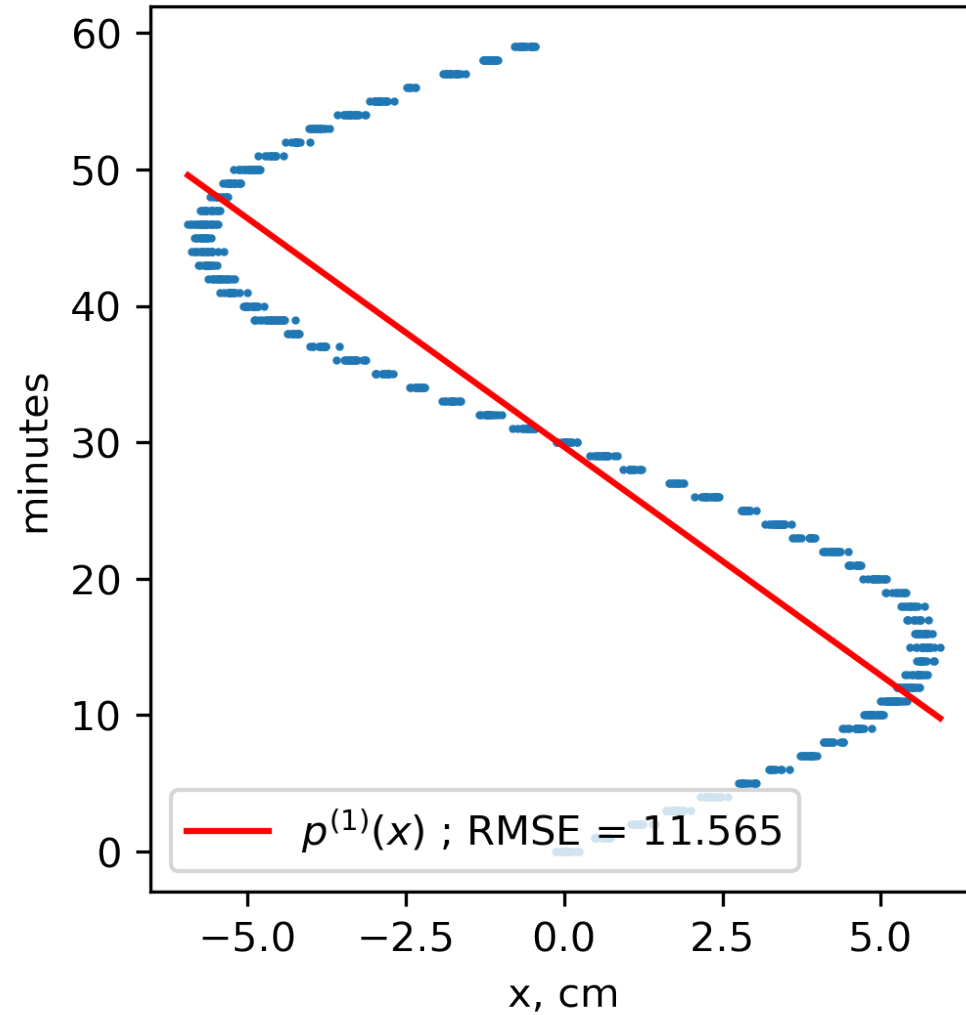
Модель в задаче восстановления регрессии: $\hat{m}_i = f(\vec{p}, x_i) = kx_i + b$

Функция потерь: $\mathcal{L}(\vec{p}, \{x_i\}, \{m_i\}) = \frac{1}{n} \sum_{i=1}^n (f(\vec{p}, x_i) - m_i)^2$

Решение (оценка параметров \vec{p}): $\vec{p}^* = \operatorname{argmin}_{\mathbb{P}} (\mathcal{L}(\vec{p}, \{x_i\}, \{m_i\}))$

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем модель посильнее

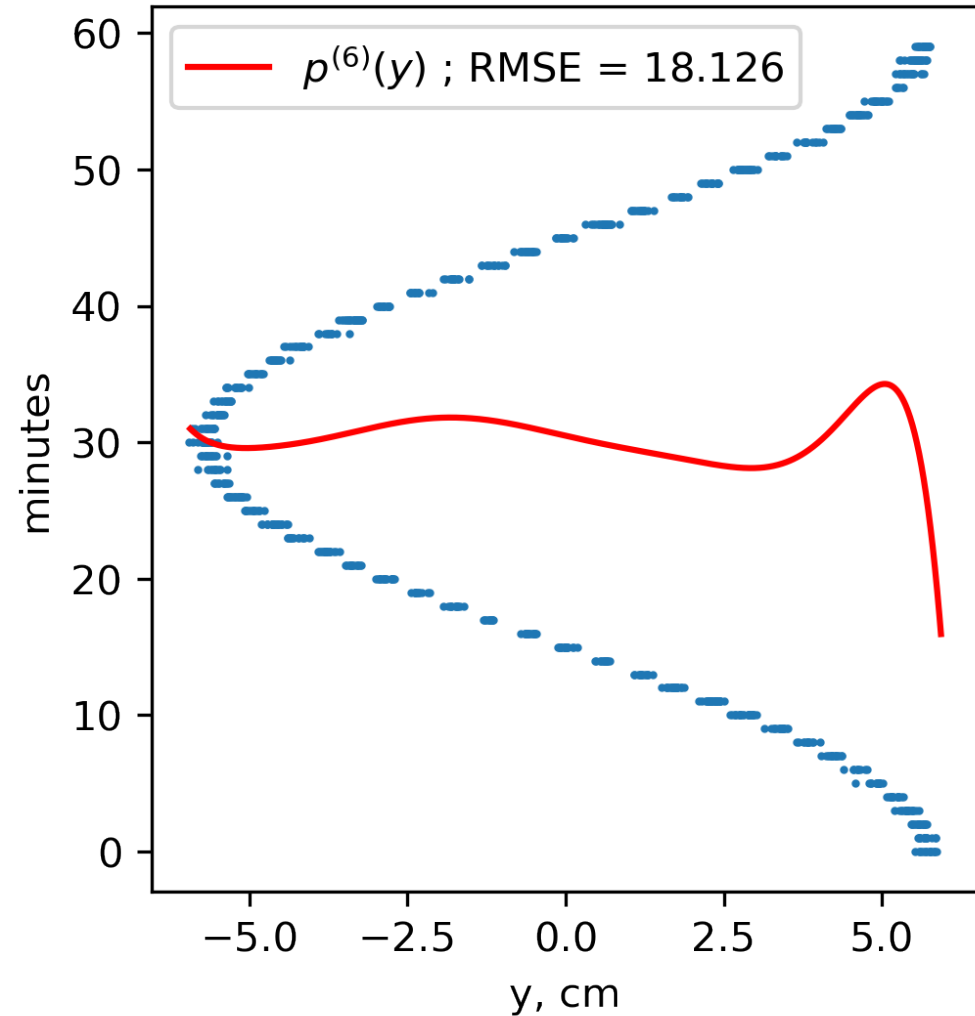
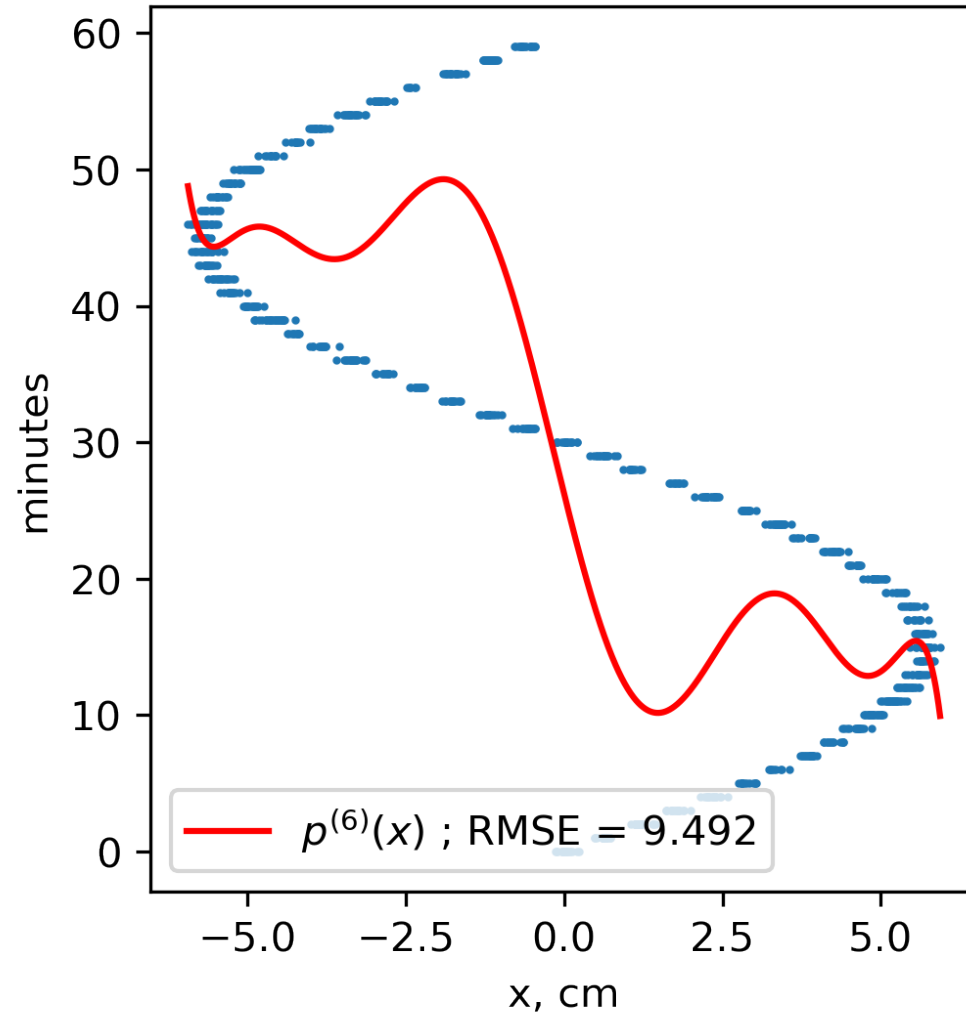
Модель в задаче восстановления регрессии: $\hat{m}_i = f(\vec{p}, \mathbf{x}_i) = \text{poly}^{(6)}(\mathbf{x}_i)$

Функция потерь: $\mathcal{L}(\vec{p}, \{\mathbf{x}_i\}, \{\mathbf{m}_i\}) = \frac{1}{n} \sum_{i=1}^n (f(\vec{p}, \mathbf{x}_i) - \mathbf{m}_i)^2$

Решение (оценка параметров \vec{p}): $\vec{p}^* = \operatorname{argmin}_{\mathbb{P}} (\mathcal{L}(\vec{p}, \{\mathbf{x}_i\}, \{\mathbf{m}_i\}))$

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем нейросеть

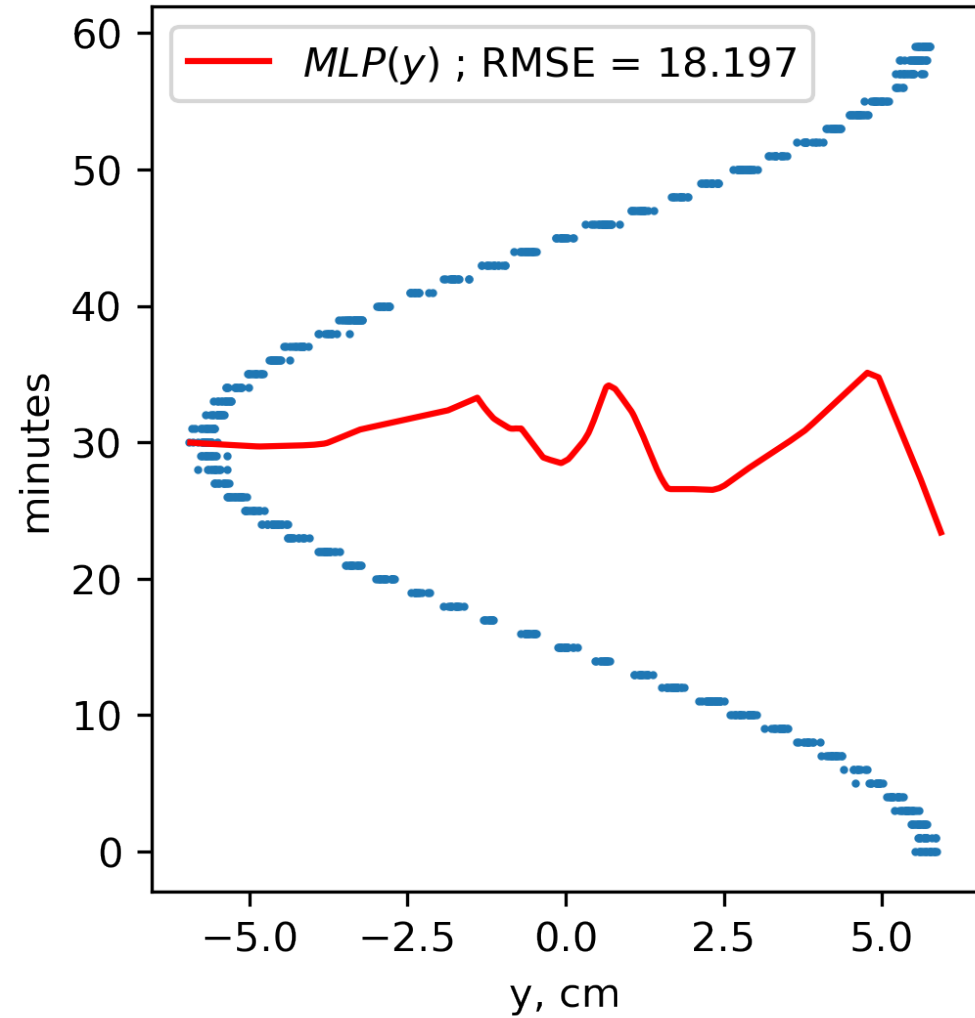
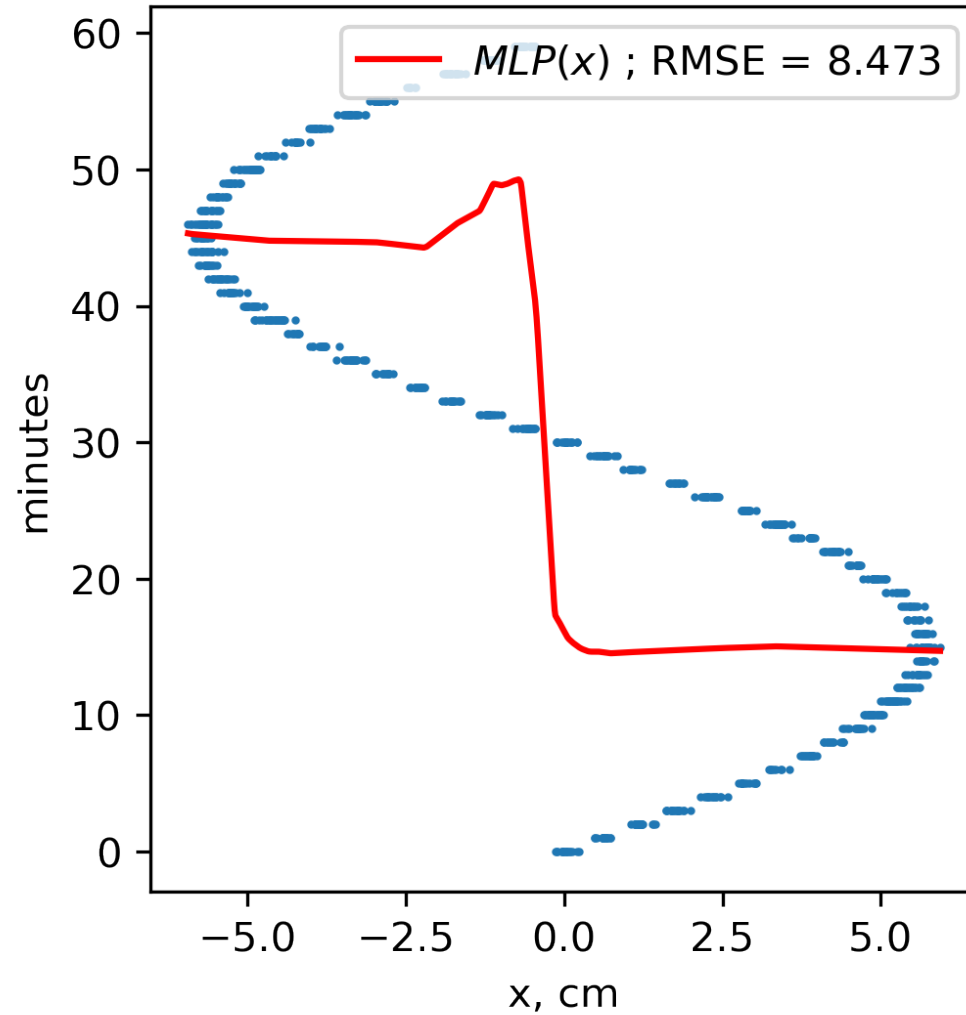
$$\hat{\mathbf{m}}_i = MLP(\vec{\mathbf{p}}, \mathbf{x}_i)$$

$$\mathcal{L}(\vec{\mathbf{p}}, \{\mathbf{x}_i\}, \{\mathbf{m}_i\}) = \frac{1}{n} \sum_{i=1}^n (MLP(\vec{\mathbf{p}}, \mathbf{x}_i) - \mathbf{m}_i)^2$$

$$\vec{\mathbf{p}}^* = \operatorname{argmin}_{\mathbb{P}} (\mathcal{L}(\vec{\mathbf{p}}, \{\mathbf{x}_i\}, \{\mathbf{m}_i\}))$$

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели



Решение задачи восстановления регрессии: ПРИМЕР

ЧТО ЖЕ ДЕЛАТЬ?

что-то не так с постановкой задачи?

(не тот тип задачи? не та целевая переменная?)

что-то не так с признаковым описанием событий?

(нерелевантное? неполное? шумное?)

что-то не так с разметкой?

(шумная? некорректная? много? мало?)

что-то не так с моделью?

(слишком простая? слишком сложная? не подходит для этой задачи?)

что-то не так с программным кодом?

что-то не так с исследователем?

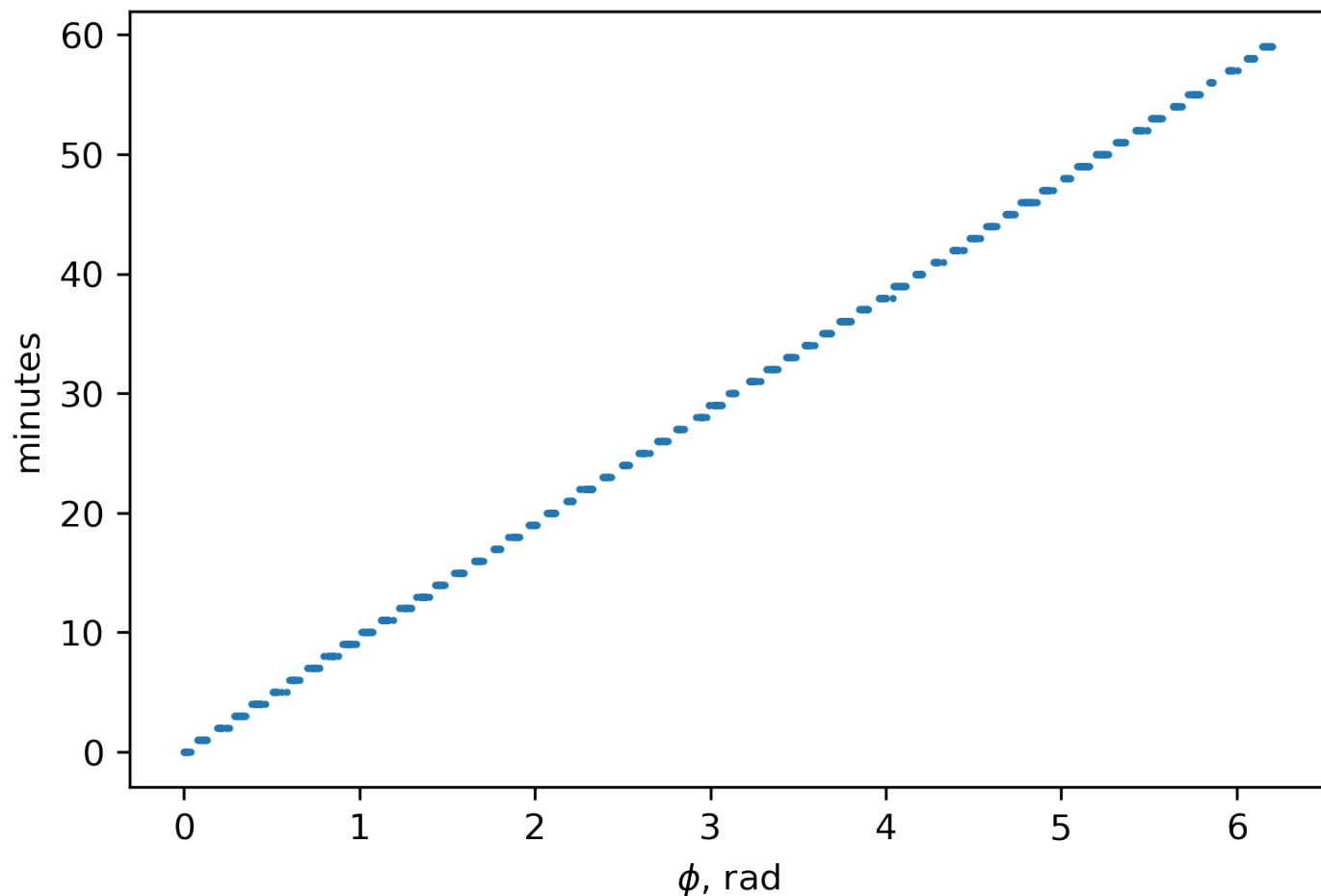
МОЖЕТ, ПРОСТО НЕТ ЗАКОНОМЕРНОСТИ?

Решение задачи восстановления регрессии: ПРИМЕР

изобретать (более информативные) признаки

Новое признаковое описание событий: \vec{x}_i - угол отклонения минутной стрелки

$$x_i = \phi = \begin{cases} \arccos\left(\frac{x}{\sqrt{x^2 + y^2}}\right), & \text{если } x \geq 0 \\ 2\pi - \arccos\left(\frac{x}{\sqrt{x^2 + y^2}}\right), & \text{если } x < 0 \end{cases}$$



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем очень слабую модель

Модель в задаче восстановления регрессии: $\hat{m}_i = f(\vec{p}, \phi_i) = k\phi_i + b$

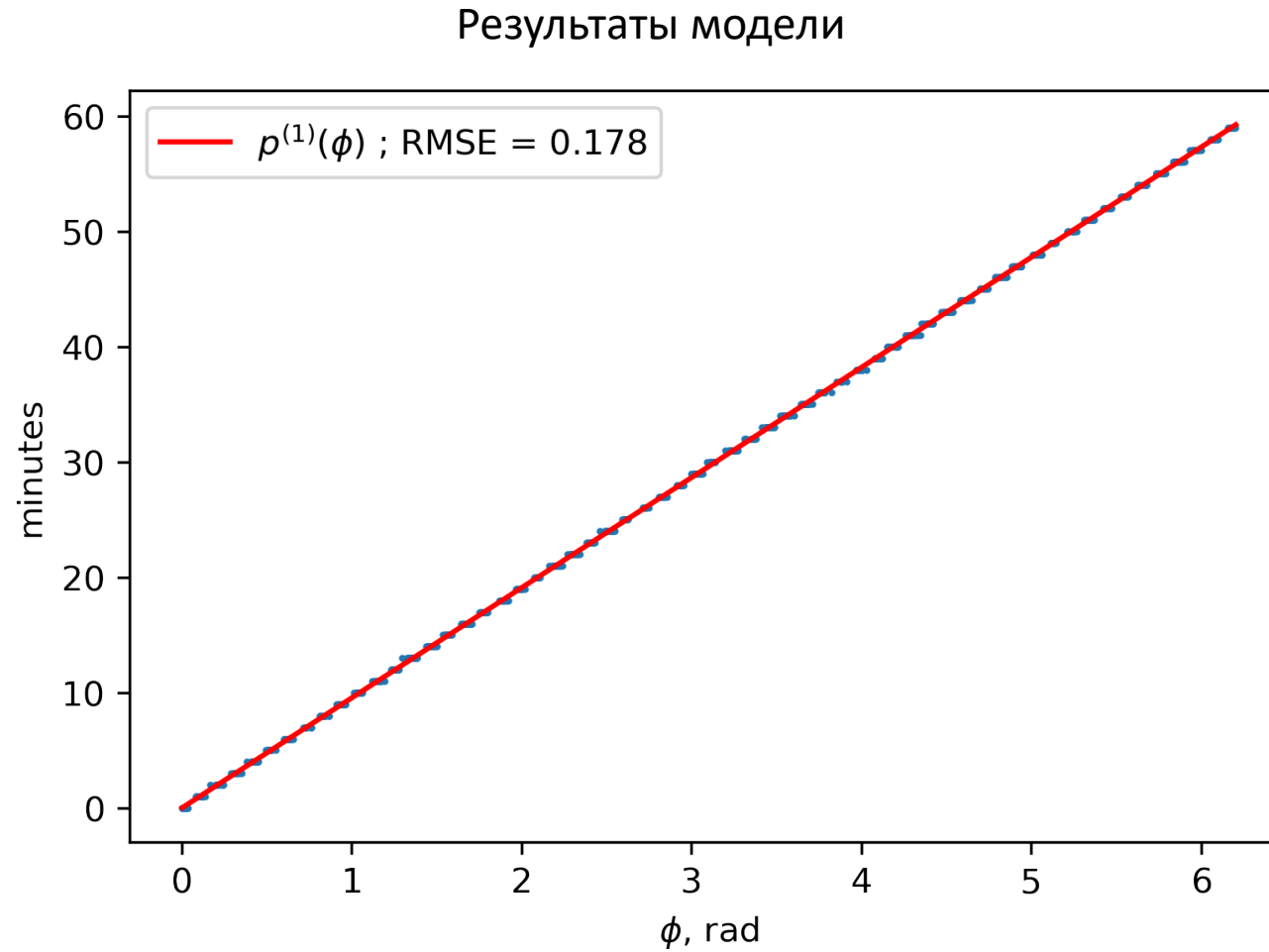
Функция потерь:

$$\mathcal{L}(\vec{p}, \{\phi_i\}, \{m_i\}) = \frac{1}{n} \sum_{i=1}^n (f(\vec{p}, \phi_i) - m_i)^2$$

Решение (оценка параметров \vec{p}):

$$\vec{p}^* = \operatorname{argmin}_{\mathbb{P}} (\mathcal{L}(\vec{p}, \{\phi_i\}, \{m_i\}))$$

Решение задачи восстановления регрессии: ПРИМЕР



Решение задачи восстановления регрессии: ПРИМЕР

использовать (более) полную информацию о событиях

Новое признаковое описание событий: \vec{x}_i - обе координаты x, y конца минутной стрелки

Возьмем нейросеть

$$\hat{m}_i = MLP(\vec{p}, \vec{x}_i)$$

$$\mathcal{L}(\vec{p}, \{\vec{x}_i\}, \{m_i\}) = \frac{1}{n} \sum_{i=1}^n (MLP(\vec{p}, \vec{x}_i) - m_i)^2$$

$$\vec{p}^* = \underset{\mathbb{P}}{\operatorname{argmin}}(\mathcal{L}(\vec{p}, \{\vec{x}_i\}, \{m_i\}))$$

Качество модели: $RMSE = 0.28m$

ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ ОБУЧЕНИЯ С УЧИТЕЛЕМ

1. формулировка задачи:

- какой тип (классификация, регрессия, другой)? Или переформулировать в легко решаемый тип!
- определиться, что есть объекты (события)
- определиться, что есть целевая переменная
- определить признаковое описание объектов (событий)
- определить критерии качества решения задачи (MSE, MAE, pattern correlation, etc.)

2. сформулировать модель:

- вид модели (линейная регрессия, дерево решений, композиционный алгоритм, нейронная сеть, etc.)
- определиться с функцией потерь (MSE, MAE, BCE, CCE, etc., комбинации)
- сложность модели (задается гиперпараметрами – настройками модели)

ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ ОБУЧЕНИЯ С УЧИТЕЛЕМ

3. подготовить данные или генератор данных:
 - стандартизировать данные (если нужно)
 - обработка пропусков, категориальных значений, кодирование текста, понижение размерности данных
 - оставить часть данных для проверки качества (train-validation-test split)
 - подготовить генератор данных с учетом стратегии скользящего контроля (cross-validation quality estimation)
4. оптимизировать модель на обучающей выборке:
 - $\hat{p} = \underset{\mathbb{P}}{\operatorname{argmin}}(L(\vec{p}, \mathcal{T}))$
5. оптимизация гиперпараметров модели и отбор моделей. Провизодится по значениям метрик качества на контрольной(контрольных) выборке(выборках)
6. оценка модели:
 - оценить качество по метрикам, определенным на этапе 1. на тестовой выборке
 - оценить неопределенность параметров модели (если возможно)
 - оценить неопределенность оценок целевой переменной

ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ ОБУЧЕНИЯ С УЧИТЕЛЕМ

6. применение модели на вновь получаемых данных:
 - оценка распределения вновь получаемых данных: генерируются ли они из того же распределения, что и обучающая выборка?
 - предобработка новых данных идентично п.3 с точностью до коэффициентов стандартизации и деталей способов предобработки
 - применение модели к предобработанным новым данным для получения значений целевой переменной
 - построение научных выводов, описание их в виде статей, получение наград в виде Нобелевских премий, etc.

Технические средства анализа данных

https://github.com/MKrinitskiy/ML4ES_2019-2020

Михаил Криницкий

krinitsky.ma@phystech.edu

К.Т.Н., Н.С.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга
климатических изменений (ЛВОАМКИ)